Thesis/
Reports
Dean, D.J.

THE NATIONAL GAP ANALYSIS PROGRAM : SENSITIVITY OF GAP ANALYSIS TO UNCERTAINTY IN FAUNAL DISTRIBUTION DATA AND HABITAT-BASED PREDICTIONS OF SPECIES OCCURRENCE

by

Denis J. Dean, et al.

The National Gap Analysis Program: Sensitivity of Gap Analysis to Uncertainty in Faunal Distribution Data and Habitat-based Predictions of Species Occurrence

A Report to
The National Council of the Paper Industry for Air and Stream Improvement

Prepared by[1]:

Denis J. Dean
Department of Forest Sciences
Colorado State University

Kenneth R. Wilson
Department of Fishery and Wildlife Biology
Colorado State University

Curtis H. Flather
USDA, Forest Service
Rocky Mountain Forest and Range Experiment Station

William C. McComb
Department of Forest Sciences
Oregon State University

April 1995

---

[1]The opinions expressed in this report are those of the authors and do not reflect the view of their respective institutions.

LIBRARY COPY
ROCKY MTN. FOREST & RANGE
EXPERIMENT STATION

Final Report for 28-C3-714 Pd'. Dean + Wilson, CSU

INTRODUCTION

In the past few years the Gap Analysis Program (GAP) has emerged as a strategy for conserving biological diversity (Scott et al. 1987, 1993). GAP studies have been conducted on a state-by-state basis, and specific guidelines for conducting GAP studies are outlined in Ma and Redmond (1992), Scott et al. (1993), Jennings (1993), and USDI, National Biological Survey (1994) (see Flather et al. 1995 for a summary of common patterns among the state study plans). In general, the basic steps in a GAP study are: (1) the development of a habitat type map based on LANDSAT Thematic Mapper (TM) imagery and other information, (2) the development of habitat-relationship models that predict where species can be found (see Morrison et al. 1992), and (3) combining these two preliminary datasets to identify and delineate areas of high species diversity. The resulting species richness map is then combined with additional maps containing land management and ownership information to identify "gaps" in the protection of species-rich areas, i.e. identify regions of high species richness that are not currently within protected areas (see Scott et al. 1993).

The most common models used in Gap Analyses to predict species occurrence are based on the intersection of coarse range map information as reflected in county-of-occurrence data, and vegetation association matrices (Scott et al. 1993, Butterfield et al. *in press*). As such, the habitat selection problem is simplified in only being concerned with predicting species presence/absence. Although simpler, defining the geographical area of a species, measuring it, and characterizing it, is controversial and very imprecise (Rapoport 1982).

Given these procedures it is obvious that GAP studies require reasonably accurate habitat type maps and sound habitat-relationship models. Unfortunately, there has been little

work on quantifying the sensitivity of GAP analyses to errors associated with these basic components (Veregrin 1989). While it is widely recognized that sensitivity analyses and validation of habitat-based predictions of species occurrence are necessary to evaluate model performance (Lyon et al. 1987, Scott et al. 1993), few species-habitat relationship data bases have been evaluated (Berry 1986). Stoms (1992) has examined the sensitivity of GAP studies to the effects of habitat type map generalization by varying minimum mapping units, but additional study aimed at exploring how habitat type errors impact GAP results remains to be done. Investigations of habitat-relationship models cast serious doubt on their overall reliability, which in turn raises concerns regarding the performance of these models in GAP studies. For example, a sensitivity analysis of a habitat-relationship model for the California condor (*Gymnogyps californianus*) indicated a relatively robust model (Stoms et al. 1992). But, Block et al. (1994) tested habitat-relationship models for relatively well studied taxa such as amphibians, reptiles, birds, and small mammals in California, and they found that agreement between predictions and residency status ranged from 48-78% for 2 databases, with errors of omission (species were not predicted but in fact exist) and commission (species were predicted but do not exist) ranging from 6-39% and 20-44%, respectively.

Since it is inconceivable that either absolutely error-free habitat type maps or perfect habitat-relationship models can be produced, all GAP studies must incorporate some amount of error from these sources. How these errors affect the results of GAP studies has not been thoroughly explored.

OBJECTIVE

The objective of this study was to investigate how uncertainty in species distributions would affect the results of a contemporary GAP study. Specifically, we investigated how both randomly occurring errors in species distribution and spatially-dependent errors in species distribution would affect the identification of species-rich areas. Note that no attempt was made to identify the cause of these errors. This study simply investigated the results of errors likely to exist in many GAP projects, regardless of source.

METHODS

The fundamental parameter being manipulated in this study was the presence or absence of a particular species in a particular habitat type polygon. If a species is actually present in a polygon and the GAP model reflects that the species is present, no error has occurred. Similarly, if a species is not present in a polygon and the model reflects that the species is not present, no error has occurred. However, if a species is actually present in a polygon but the model indicates that the species is not present, an error of omission occurs. An error of commission occurs if, in fact, a species is not present in a polygon, but the model indicates that the species is present.

This study adopted GAP data received from the Oregon Department of Fish and Wildlife as a baseline from which we estimated the effects of introduced errors. The Oregon data consisted of a digital habitat type map that divided the entire state into over 12,000 habitat polygons based on vegetative cover types (which were identified through the processing of satellite imagery), physiographic province (based on Puchy and Marshall 1993), and political

boundaries. The Oregon data also included habitat-relationship models for over 800 wildlife species that associated species occurrence based on vegetative cover, physiographic region, and political subdivision data layers. By combining these datasets it was possible to construct a species distribution map showing the presence or absence of each species in each habitat polygon.

Monte Carlo techniques were used to simulate the effects of error on species distribution maps. Each iteration of the Monte Carlo simulation randomly introduced a predefined amount of error into the species distribution map, thereby producing a new error-filled species distribution map. The areas of high species richness identified on this error-filled map were then compared to the high richness areas identified in the original map to determine what impacts the introduced errors had on the identification of areas of high species richness.

Two different techniques were used to introduce error into the species distribution map. The first technique introduced *spatially independent* errors throughout the vegetation map without regard to the range of any particular wildlife species. Thus, any polygon in the habitat map where species *X* did in fact occur was a candidate for an error of omission of species *X* and any polygon where species *X* did not in fact occur was a candidate for an error of commission of species *X*. Errors of this sort are likely to be present in GAP studies due to the inaccuracies of habitat mapping--errors in such maps are unlikely to be dependent upon the ranges of the wildlife species found in the mapped region.

The second technique introduced *spatially dependent* errors along the edges of each species "true" range, where the true range was assumed to be the species' range as determined

by the original Oregon data. Thus, spatially dependent errors could only occur along the edges of the range of the species, with errors of omission occurring in edge polygons within the true range of the species and errors of commission occurring in polygons just outside the true range. Errors of this sort are likely to occur in GAP studies if inaccuracies in species predictions from habitat-relationship models occurs at the boundary of a species' range, or if there is difficulty associated with precisely defining the boundaries of habitat polygons.

GAP results were measured based on the characteristics and distributions of areas of high species diversity. A total of 12 hypotheses were evaluated in this study based on areas of high species richness. High species richness was evaluated for 3 measures (total area, perimeter, and number of disjoint or noncontiguous polygons) and two different types of errors (spatially dependent and spatially independent) at introduced error rates of 5% and 20%. Each hypothesis was similar to the following:

H_0: Spatially independent errors have no significant impact on the total area of regions of high species richness present in the results of a GAP analyses.

The remaining 11 hypothesis simply enumerated all possible combinations of error type and measure of GAP outcomes. Each of the hypotheses were tested using a t-test ($\alpha = 0.05$) by comparing the average changes in area, perimeter, and number of disjoint polygons from the Monte Carlo simulations to zero, the expected value of the change under the null hypotheses.

The Monte Carlo simulations were extremely time consuming and required large amounts of computer storage space, thus we decided to use only a subset of the original data in the analysis. A smaller vegetation map was extracted from the original data and a subset of

the 800 available species were chosen for analysis. The smaller map consisted of 3 of the 12 physiographic provinces (East Slope Cascades, Basin and Range, and Owyhee Uplands) from the original map and contained 3,238 habitat polygons. The selected species subset consisted of ten rodents: mountain beaver (*Aplodontia rufa*), deer mouse (*Peromyscus maniculatus*), yellow-pine chipmunk (*Tamias amoenus*), white-tailed antelope squirrel (*Ammospermophilus leucurus*), northern flying squirrel (*Glaucomys sabrinus*), California kangaroo rat (*Dipodomys californicus*), desert woodrat (*Neotoma lepida*), heather vole (*Phenacomys intermedius*), water vole (*Micotus richardsoni*) and sagebrush vole (*Lemmiscus curtatus*). These reduced datasets were joined to create a digital map showing the distributions of each of the ten rodents throughout the study region which then became the basis for this study's Monte Carlo analyses (Figure 1).

Each iteration of the Monte Carlo simulation introduced a predefined level of either spatially dependent or spatially independent error into the original data. Error amounts were measured as percentages. A 5% level of commission error for species *X* meant that each polygon where species *X* did not occur had an 5% chance of having species *X* introduced. Similarly, an 5% level of omission error for species *X* meant that each polygon where species *X* occurred had an 5% chance of having species *X* removed. Errors of omission and commission were simulated in much the same fashion for both the spatially independent and spatially dependent simulations, but, in the spatially dependent simulations, errors could only take place in the appropriate regions along the edges of the various species' ranges whereas spatially independent errors could take place anywhere throughout the study area. The appropriate regions for spatially dependent error of omission for species *X* were found by

finding the total area comprising the species' range and contracting the boundary of the area until only 90% of the original area remained. The region between the original boundary and the 90% area boundary were deemed to be suitable for errors of omission. The appropriate regions for spatially dependent errors of commission were then found by expanding the original species *X* range boundary by the same amount as it was decreased in the previous step and deeming the region between the original and expanded boundaries the appropriate region for commission errors.

Two error levels (5 and 20%) for both omission and commission errors were simulated. These two rates were used for both the spatially independent and the spatially dependent cases resulting in a total of four Monte Carlo simulations. In all cases each simulation consisted of 50 iterations, and each iteration produced a map showing places of high species richness (which was arbitrarily defined as places where 5 or more species occurred).

The areas, perimeters, and number of disjoint polygons making up areas of high species richness were recorded after each iteration. These recorded outputs were then compared to the area, perimeter, and number of disjoint polygon measures taken from the unaltered dataset using simple comparison of means procedures (t-tests).

RESULTS

The original base map for the 3 physiographic regions included approximately 89,195 km^2. Approximately 82,904 km^2 (93%) of the base map had low species diversity ($\leq$ 4 species) and 6,291 km^2 (7%) had high species diversity ($\geq$ 5 species) (see Figure 2).

Spatially Independent Error: The null hypothesis was rejected for the 6 cases with spatially independent error indicating that there was a significant change in species rich areas based on the 3 variables measured (average changes in area, perimeter, and number of disjoint polygons) (Table 1). With a 5% error, species rich areas increased by 150% over the base map, from 6291 km^2 to 15774 km^2. With a 20% error, species rich areas increased by nearly 400% over the base map, from 6291 km^2 to 311987 km^2. The number of disjoint polygons increased from 52 for the base map to 224 for the 5% error simulations and from 52 to 313 for the 20% error simulations. Overall, a large portion of the area outside the original species rich area has the potential for high species richness when even a small amount (5%) of spatially independent error is simulated, and this amount increases as error increases (20%) (Figure 2).

Spatially Dependent Error: The null hypothesis was also rejected for the 6 cases with spatially dependent error resulting in a significant change in species richness based on the 3 variables measured (Table 2). Compared to the area of species richness for the base map (6291 km^2), the average size of the species rich area declined to 6173 km^2 (1.8 % decrease) and 5863 km^2 (6.8 % decrease) for 5% and 20% spatially dependent error, respectively. The total perimeter increased by 1522 km and 4196 km for the 5% and 20% independent spatial errors, respectively. The number of disjoint polygons increased from 52 for the base map to 142 for the 5% error simulations and to 439 for the 20% error simulations.

On the surface, the map format of the spatially dependent error simulation appears to show no visible error in the GAP results due to either the 5% or 20% introduced error (Figure 3). The discrepancy between the results in Figure 3 and the significance of the test results in Table 2 (which strongly support the hypothesis that even small amounts of spatially-dependent

error cause significant changes in the ability of GAP studies to reliably identify areas of high species diversity) is an artifact of map scale. Errors are present in Figure 3, but they are not visible at the current map scale. If the scale is increased, then errors become increasingly visible as regions of error become apparent (Figure 4). Errors of omission are shown in Figure 4 as those places within the region of true high species richness (the region of true high species richness is outlined in red in Figure 4) that are shaded in some color other than white--these are areas that at least some of the Monte Carlo iterations failed to identify as having high species richness. Errors of commission are shown in Figure 4 as places outside the region of true high species richness that are shaded is some color other than black--at least some iterations of the Monte Carlo simulator erroneously identified these areas as having high species richness. By arbitrarily increasing the scale of the map we can make these shaded areas arbitrarily more obvious, so the fact that they were not obvious in the original Figure 2 was simply an artifact of the map scale.

DISCUSSION

The 2 types of errors simulated in this study can be viewed as a continuum with each at extreme ends. A priori, it is reasonable to expect that our method of simulating spatially independent error throughout the vegetation map without regard to a species range should have the greatest effect on determining areas of high species richness, because errors of omission and commission are allowed to enter in areas where animals might obviously not exist. Alternatively, the method for simulating spatially dependent error should result in fewer changes in the areas associated with high species richness, because our method assumes that

animals have some definable range and errors in omission and commission can only occur within a restricted area of the range boundary. Spatially dependent error is also conservative because the $\pm 10\%$ error used in this study (as defined on page 7) is of a finer scale than the finest scale of error that might be produced in a GAP-based prediction of species occurrence at the habitat type/physiographic region/county polygon level.

The most realistic type of error that exists in a GAP dataset is probably a combination of independent and dependent errors. For example, even after an animal's range has been identified, there can be independent errors that occur within the species' range. This would amount to simulating independent or random error only within the known range of the species. This type of error might occur because of random errors associated with vegetation mapping (e.g. using remote sensing) (Khagendra and Bossler 1992). In addition, there is bound to be dependent error associated with the wildlife habitat relationship model (see Morrison et al. 1992, Block et al. 1994) or the fact that the wildlife data is compiled from data collected from a variety of sources (see Scott et al. 1993).

In our simulations, spatially independent error resulted in large increases in species rich areas. Failure to recognize this type of error could result in extremely unreliable information. Although, spatially dependent error resulted in a significant decline in species rich areas the amount (2-7% depending on error level) was small relative to the size of the base map area. The area subject to change under the spatially dependent error simulations was approximately 1300 km^2 or $\pm 10\%$ of the 6291 km^2 area of the original species rich area. This amounts to a 9 to 33% change in area when focusing exclusively on the 1300 km^2 affected area. Even if the GAP study is used to make a preliminary determination of species rich areas as suggested by

Scott et al. (1993), the implications are that not considering spatially independent and dependent errors could result in focusing the preliminary study on a large number of incorrect areas.

The error rates for omission and commission found in the Block et al. (1994) study suggest that the 5 and 20% error rates used in this study are not unreasonable. Both types of errors might directly be attributable to the wildlife habitat relationship models or from inaccuracies in the species database due to poor data or insufficient sampling. In addition, simple habitat relationship models do not account for community and ecosystem processes that may be important in determining species occurrences.

An alternative to using species richness, termed "set-coverage", has been proposed for GAP (Scott et al., *in prep.*, A. R. Kiester, U.S. For. Serv., *pers. comm.*, Wright et al. 1994). Set-coverage defines the minimum number of areas needed to ensure that all species are represented (Margules et al. 1988). This approach may lead to a more comprehensive representation of species within a state, but it will probably do nothing to mitigate the effects of errors in omission and commission. Quite the contrary, a method such as set coverage may be more vulnerable to such errors because the approach may more closely emulate spatially independent error.

Many authors have argued for the incorporation of error modeling into GIS (Chrisman 1989, Veregin 1989, Lanter and Veregin 1992). Without some indication of the sensitivity of GAP to the types of errors evident in GIS databases, choosing areas based on criteria such as species richness will prove difficult. Incorporation of error modeling or sensitivity analysis

capabilities into a GAP seems essential, otherwise users of GAP will be unaware of the uncertainty associated with their analysis.

ACKNOWLEDGMENTS

We would like to thank T. O'Neill for providing the Oregon State GAP, and K. G. Croteau for programming help.

LITERATURE CITED

Berry, K. H. 1986. Introduction: development, testing, and application of wildlife-habitat models. Pages 3-4 *in* J. Verner, M. L. Morrison, and C. J. Ralph, eds. Wildlife 2000: modeling habitat relationships of terrestrial vertebrates. Univ. Wisconsin Press, Madison.

Block, W.M., M.L. Morrison, J. Verner, and P.N. Manley. 1994. Assessing wildlife-habitat-relationships models: a case study with California oak woodlands. Wildlife Society Bulletin 22:549-561.

Butterfield, B. R., B. Csuti, and J. M. Scott. *in press*. Modeling vertebrate distributions for Gap Analysis. *in* R. I. Miller, ed. Mapping the diversity of nature. Chapman and Hall, New York, N.Y.

Chrisman, N. R. 1989. Modeling error in overlaid categorical maps. Pages 21-34 *in* M. Goodchild, and S. Gopal, eds. The accuracy of spatial databases. Taylor and Francis, New York, N.Y.

Flather, C. H., D. J. Dean, K. R. Wilson, and W. C. McComb. 1995. The National Gap Analysis Program: A review and inspection of ecological assumptions. Final report to National Council of the Paper Industry for Air and Stream Improvement.

Jennings, M. D. 1993. Natural terrestrial cover classification: assumptions and definitions. Gap Analysis Tech. Bull. 2. U.S. Fish and Wildlife Service, Idaho Cooperative Fish and Wildlife Research Unit, Moscow. 28pp.

Khagendra, T., and J. Bossler. 1992. Accuracy of spatial data used in geographic information systems. Photogr. Eng. and Rem. Sens. 58:835-841.

Lanter, D. P., and H. Veregin. 1992. A research paradigm for propagating error in layer-based GIS. Photogr. Eng. and Rem. Sens. 58:825-833.

Lyon, J. G., J. T. Heinen, R. A. Mead, and N. E. G. Roller. 1987. Spatial data for modeling wildlife habitat. J. Surv. Engineer. 113:88-100.

Ma, Z., and R. L. Redmond. 1992. Building attribute tables for raster GIS files with ARC/INFO. Gap Analysis Tech. Bull. National GAP Analysis Research Project, U.S. Fish and Wildl. Serv., Idaho Coop. Fish and Wildl. Res. Unit, Univ. Idaho, Moscow. 2pp.

Margules, C. R., A. O. Nicholls, and R. L. Pressey. 1988. Selecting network reserves to maximise biological diversity. Biol. Conserv. 43:63-76.

Morrison, M. L., B. G. Marcot, and R. W. Mannan. 1992. Wildlife-habitat relationships: concepts and applications. Univ. Wisconsin Press, Madison. 343pp.

Puchy, C. A. and D. B. Marshall. 1993. Oregon wildlife diversity plan, 2nd ed. Oregon Fish and Wildlife Commission, and Oregon Department of Fish and Wildlife.

Rapoport, E. H. 1982. Areography: geographical strategies of species. Pergamon Press, New York, N.Y. 269pp.

Scott, J. M., B. Csuti, J. D. Jacobi, J. E. Estes. 1987. Species richness: a geographic approach to protecting future biological diversity. BioScience 37:782-788.

Scott, J. M., F. Davis, B. Csuti, R. Noss, B. Butterfield, C. Groves, H. Anderson, S. Caicco, F. D'Erchia, T. C. Edwards, Jr., J. Ulliman, and G. Wright. 1993. Gap Analysis: a geographic approach to protection of biological diversity. Wildl. Monogr. 123. 41pp.

Scott, J. M., A. R. Kiester, B. Csuti, R. F. Noss, and B. Butterfield. *in prep.* Gap analysis: a spatial approach to identifying representative areas for maintaining biodiversity.

Stoms, D. M. 1992. Effects of habitat map generalization in biodiversity assessment. Photogr. Eng. and Rem. Sens. 58:1587-1591.

Stoms, D. M., F. W. Davis, and C. B. Cogan. 1992. Sensitivity of wildlife habitat models to uncertainties in GIS data. Photogr. Eng. and Rem. Sens. 58:843-850.

U.S. Department of the Interior, National Biological Survey. 1994. A handbook for gap analysis: a geographical approach to planning for biological diversity. U.S. Department of the Interior, National Biological Survey, Washington, D.C., unnumbered.

Veregin, H. 1989. Error modeling for the map overlay operation. Pages 3-18 *in* M. Goodchild, and S. Gopal, eds. The accuracy of spatial databases. Taylor and Francis, New York, N.Y.

Wright, R. G., J. G. MacCracken, and J. Hall. 1994. An ecological evaluation of proposed new conservation areas in Idaho: evaluating proposed Idaho National Parks. Conserv. Biol. 7:207-216.

Table 1. Comparison of the change in species richness (greater than or equal to 5 species per polygon) for 50 repetitions as measured by the difference in the average area (A), perimeter (P), and number of noncontiguous regions (NC) between the base map and the map with independent spatial errors of 5 and 20 %. All comparisons were significant at p = 0.001 using a t-test comparing the mean change to zero.1

	5% error		20% error	
Variable	Mean Change	SD	Mean Change	SD
Area (km^2)	9483.11	1841.00	24902.80	2514.50
Perimeter (km)	8841.69	986.74	18510.08	751.05
NC	172.56	13.02	261.00	29.77

Table 2.Comparison of the change in species richness (greater than or equal to 5 species per polygon) for 50 repetitions as measured by the difference in the average area (A), perimeter (P), and number of noncontiguous regions (NC) between the base map and the map with dependent spatial errors of 5 and 20 %. All comparisons were significant at p = 0.001 using a t-test comparing the mean change to zero.

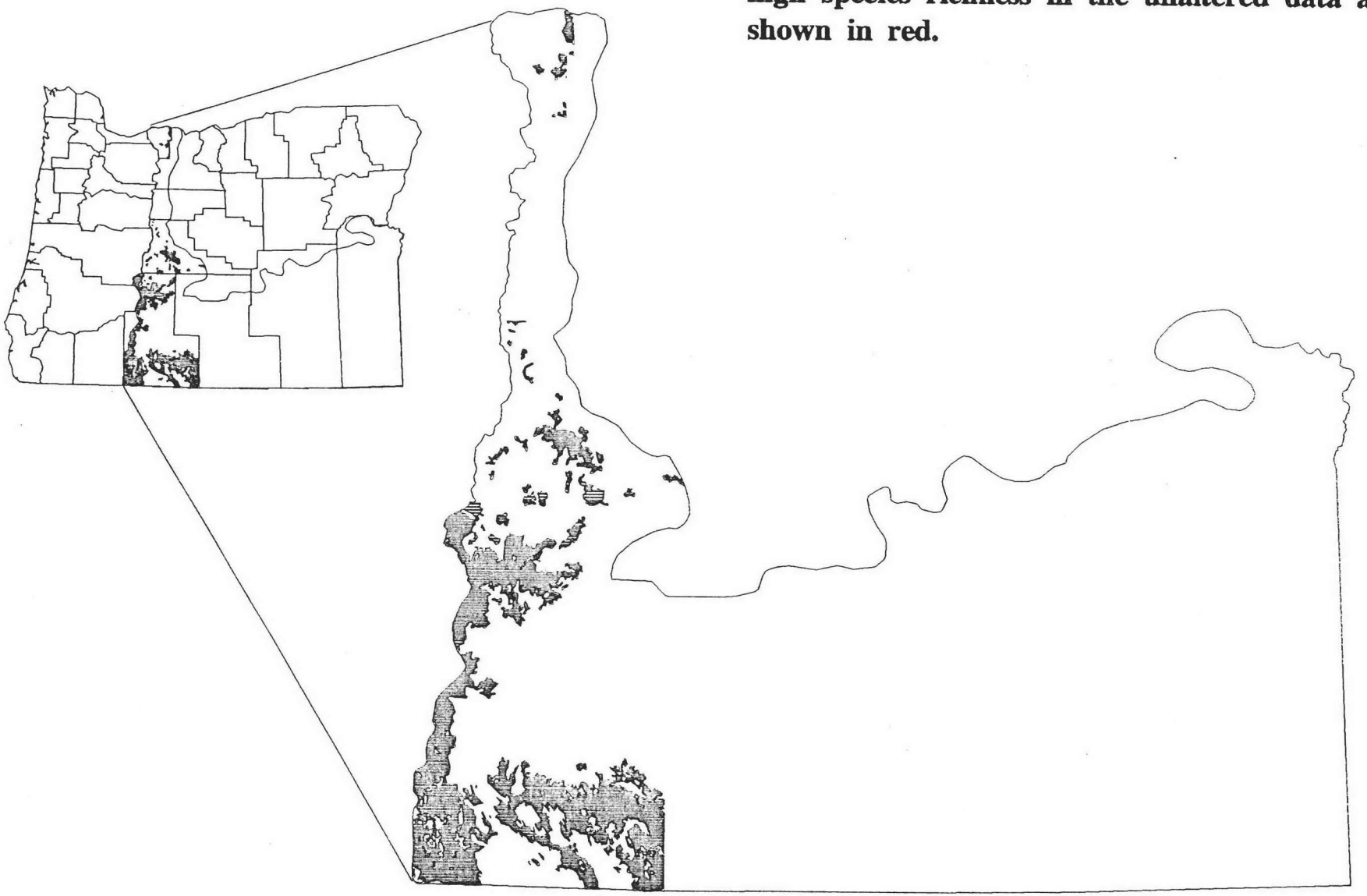
	5% error		20% error	
Variable	Change	SD	Change	SD
Area (km^2)	-117.70	28.31	-427.74	52.59
Perimeter (km)	1522.47	281.72	4195.96	312.26
NC	89.72	11.08	387.20	26.89

Figure 1. Study area and regions of high species diversity. These data constituted the basis of the Monte Carlo analysis.

Figure 2. Results of the spatially independent error analysis. Areas outlined in red have high species diversity (5 or more species) in the original, unaltered Oregon data. Gray shades indicate the number of times the Monte Carlo simulator identified an area as have high species diversity; areas shown in light gray had high species diversity in most or all of the simulations while areas shown in dark gray or black seldom if ever had high species diversity.

Figure 3. Results of the spatially dependent error analysis. Areas outlined in red have high species diversity (5 or more species) in the original, unaltered data. Gray shades indicate the number of times the Monte Carlo simulator identified an area as having high species diversity; areas shown in light gray had high species diversity in most or all of the simulations while areas shown in dark gray or black seldom if ever had high species diversity.

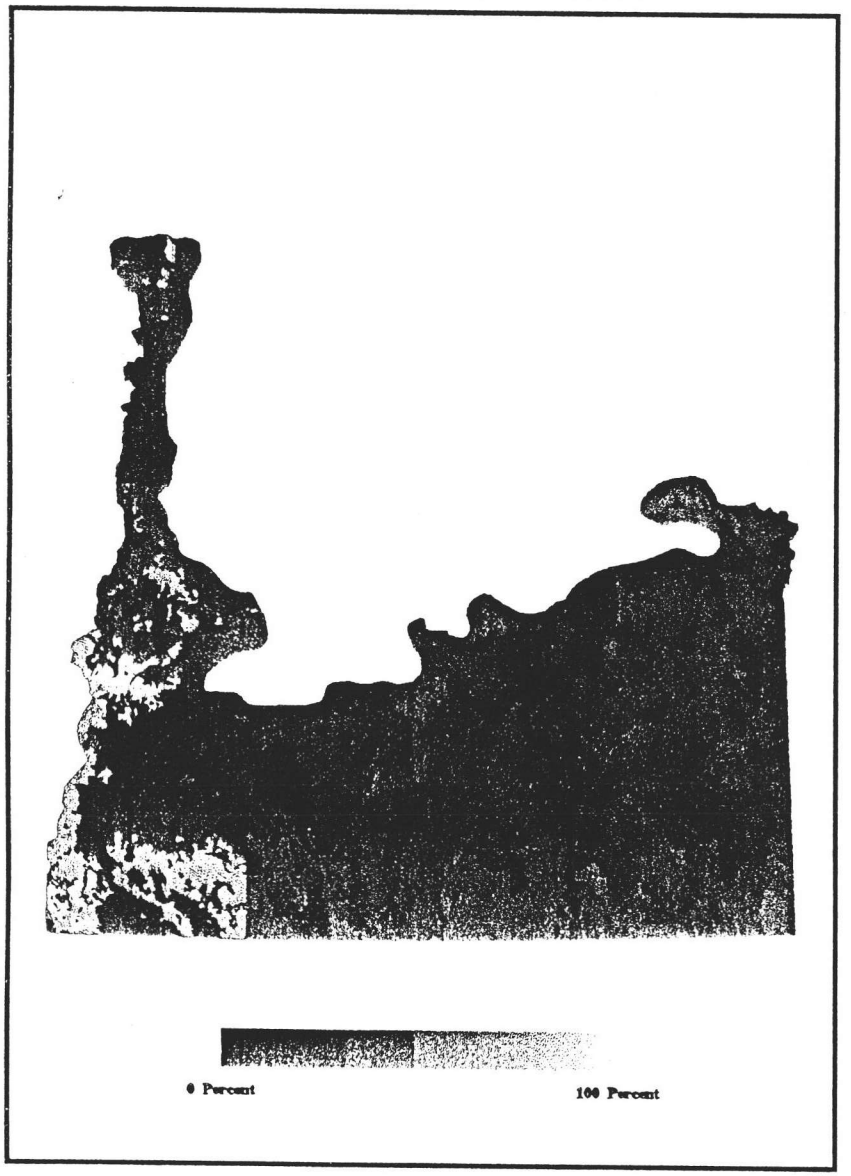
Figure 4. Detailed results of the spatially dependent error analysis. This map represents a close-up of a small portion of the map shown in Figure 2. Areas outlined in red have high species diversity (5 or more species) in the original, unaltered Oregon data. Gray shades indicate the number of times the Monte Carlo simulator identified an area as have high species diversity; areas shown in light gray had high species diversity in most or all of the simulations while areas shown in dark gray or black seldom if ever had high species diversity.



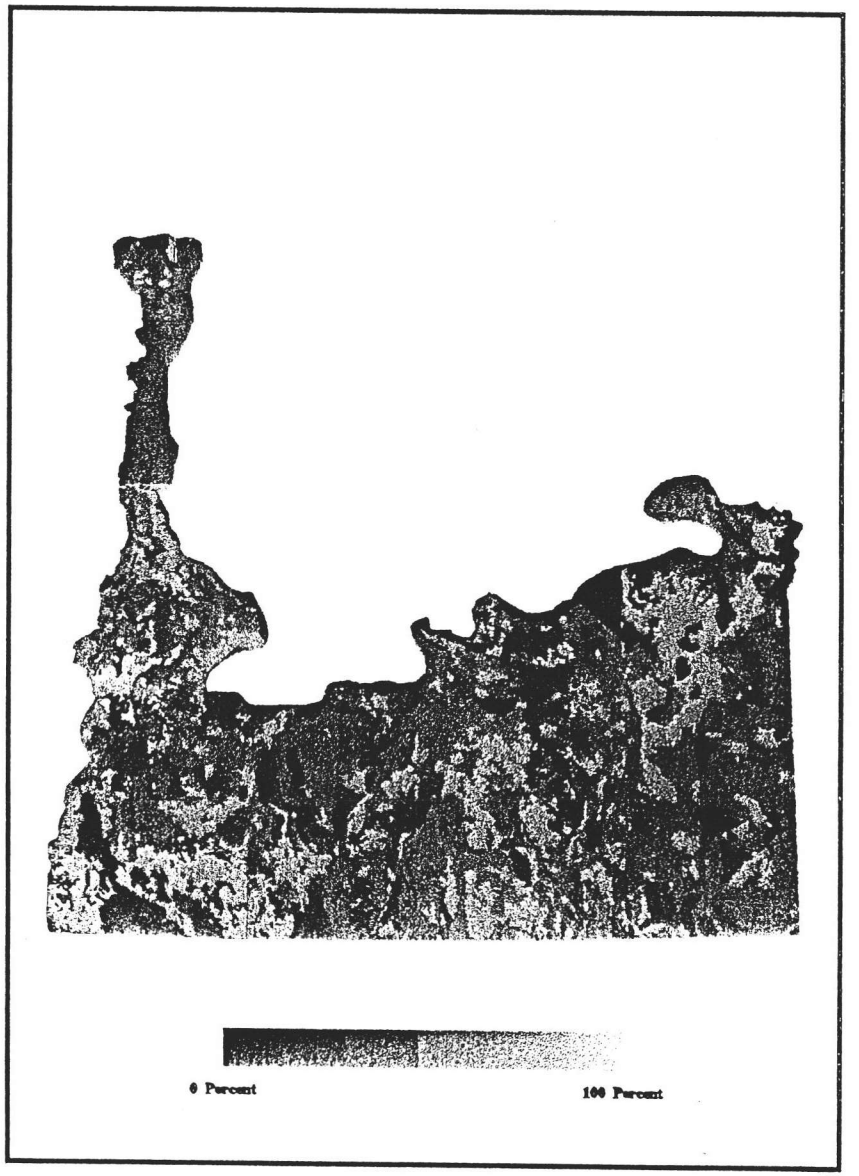
Study area used in Monte Carlo analysis. Areas with high species richness in the unaltered data are shown in red.

SPATIALLY INDEPENDENT ERROR SIMULATION

5% Error of Ommission
5% Error of Commission



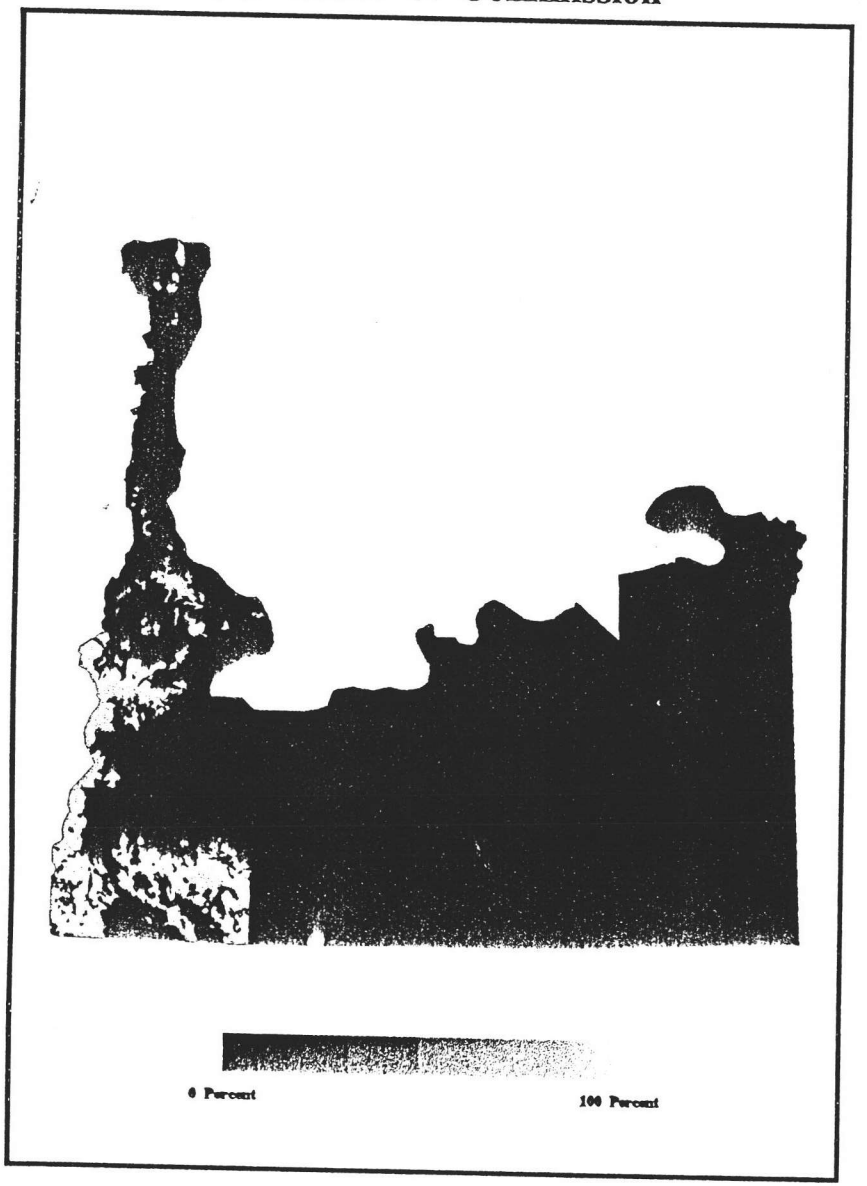
20% Error of Ommission
20% Error of Commission



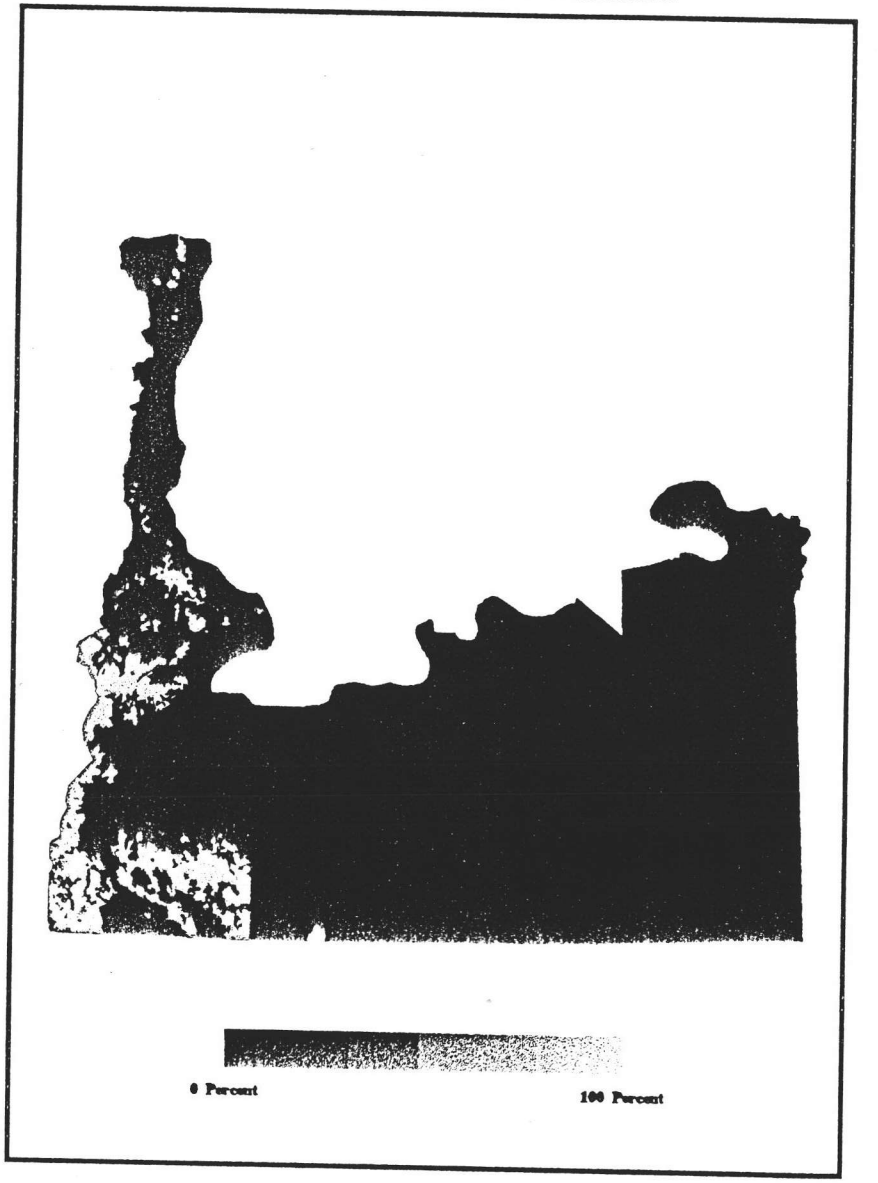
Shades indicate percentage of simulation runs where areas were found to have high species diversity (5 or more ~~species~~ from a total of 10 species considered). Outlined areas show location with high species diversity in the original dataset.

SPATIALLY DEPENDENT ERROR SIMULATION

5% Error of Ommission
5% Error of Commission



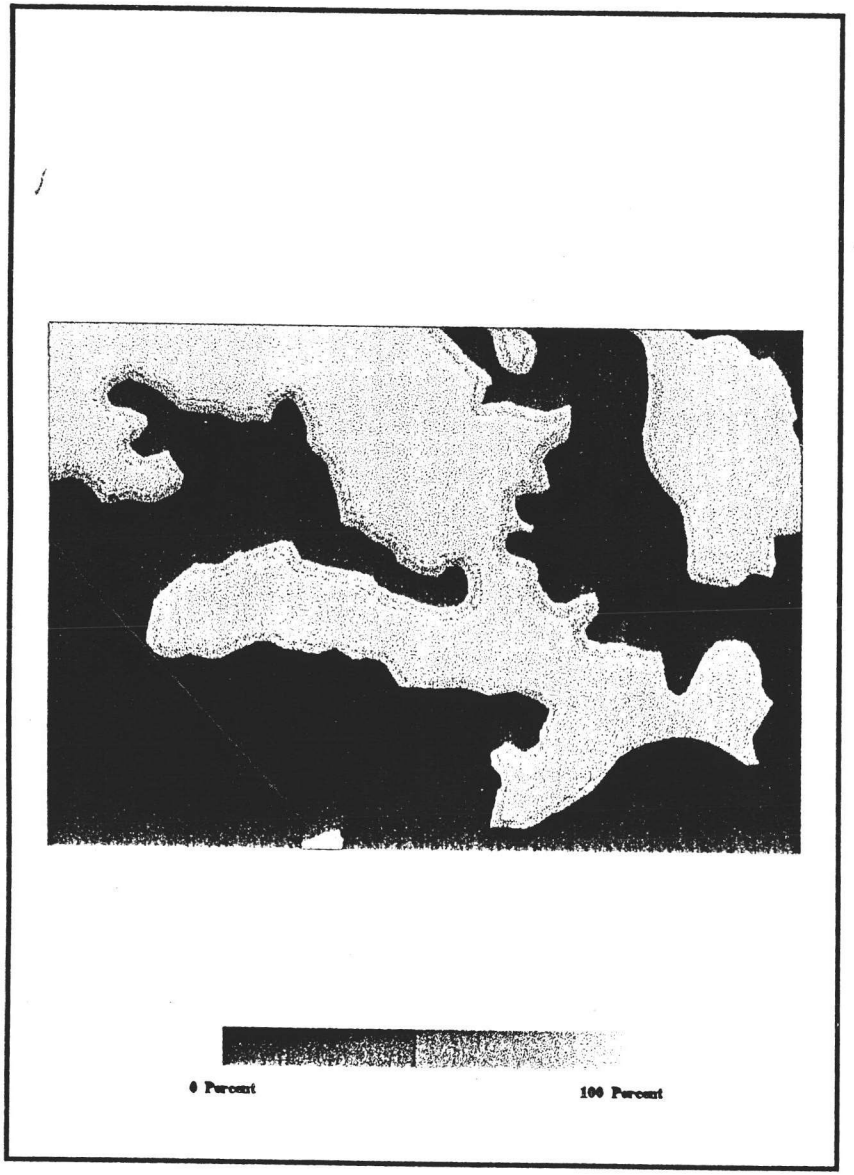
20% Error of Ommission
20% Error of Commission



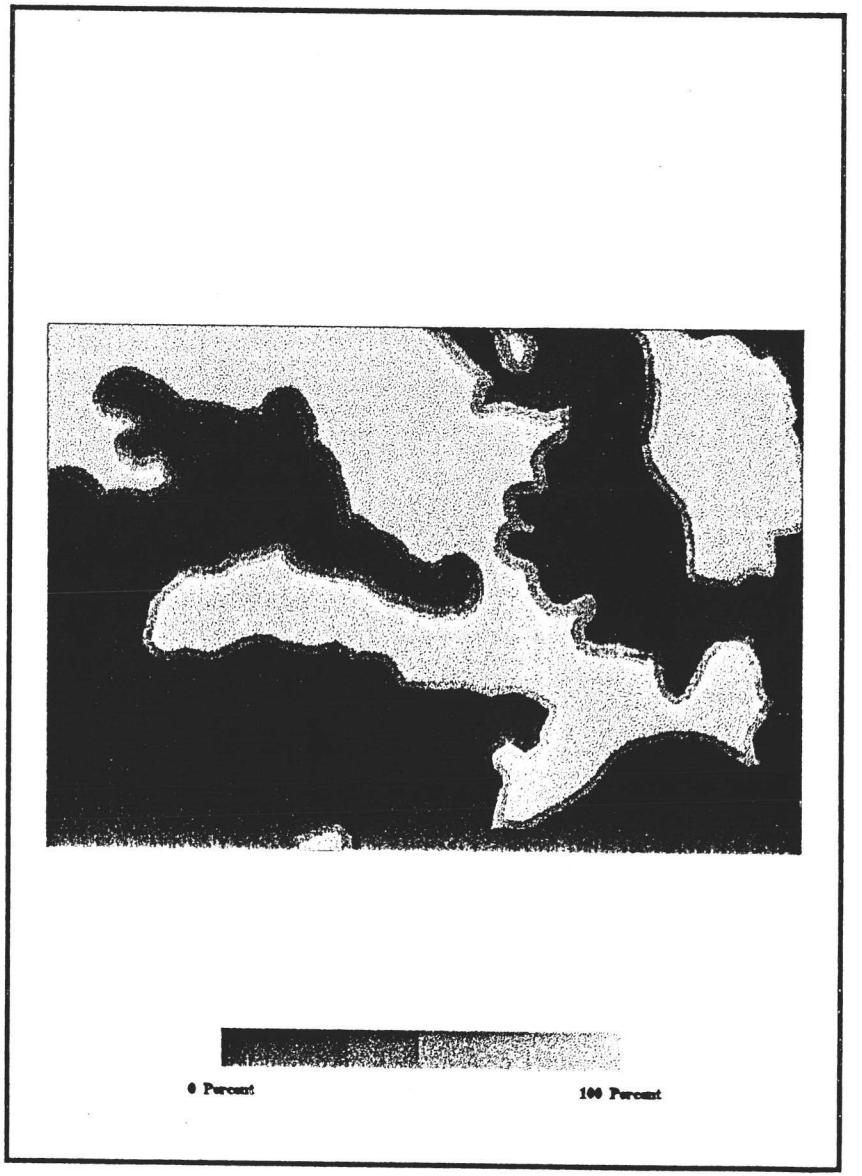
Shades indicate percentage of simulation runs where areas were found to have high species diversity (5 or more ~~species~~ from a total of 10 species considered). Outlined areas show location with high species diversity in the original dataset.

SPATIALLY DEPENDENT ERROR SIMULATION

5% Error of Ommission
5% Error of Commission



20% Error of Ommission
20% Error of Commission



Shades indicate percentage of simulation runs where areas were found to have high species diversity (5 or more present from a total of 10 species considered). Outlined areas show location with high species diversity in the original dataset.